JOURNAL OF MULTIVARIATE ANALYSIS 12, 248–255 (1982)

Cramér–von Mises Type Estimation of the Regression Parameter: The Rank Analogue

MARK A. WILLIAMSON

Michigan State University

Communicated by T. L. Lai

A point estimator based on minimization of the rank analogue of the Cramér–von Mises statistic is proposed for the slope parameter β in the simple linear regression model. The asymptotic distribution of the estimator is derived and its variance is compared to the asymptotic variances of several common estimators for β at various underlying distributions.

0. INTRODUCTION

Consider the simple linear regression model $X_{ni} = \beta_0 + \beta c_{ni} + \varepsilon_{ni}$, $1 \leqslant i \leqslant n$, where $c_{n1} \leqslant c_{n2} \leqslant \cdots \leqslant c_{nn}$ are known constants, not all equal, β_0 and β are unknown parameters, and the ε_{ni} are iid F for F a distribution with a continuous bounded density f satisfying $f(x) > 0$ a.e. on $\{x: 0 < F(x) < 1\}$. We regard β_0 as a nuisance parameter and seek to estimate β.

As given by Hájek and Sidák [2, p. 103], the rank analogue of the Cramér–von Mises test for $H_0: \beta = \Delta$, Δ a known constant, is based on the statistic

$$M(\Delta) = \int_0^1 \left[\sum_{i=1}^{n} (c_{ni} - \bar{c}_n)\, I(R_{ni\Delta} \leqslant nt) \right]^2 dt,$$

where $\bar{c}_n = n^{-1} \sum_{i=1}^{n} c_{ni}$ and $R_{ni\Delta}$ is the rank of $X_{ni} - \Delta c_{ni}$ among $\{X_{nj} - \Delta c_{nj};\ 1 \leqslant j \leqslant n\}$. Since H_0 is rejected only for large values of $M(\Delta)$, it seems reasonable to define an estimator $\hat{\beta}$ for β based on the minimization of M in Δ. The definition and properties of such an estimator are discussed in the sections that follow.

Received April 6, 1979; revised August 3, 1981.
AMS 1980 subject classifications: Primary 62G05; Secondary 62G35.
Key words and phrases: Cramér–von Mises estimation, regression parameter.

0047-259X/82/020248–08$02.00/0

1. Notation and Preliminaries

To the assumptions of the model introduced in Section 0 we add the following:

$$\lim_{n} \frac{\sum (c_{ni} - \bar{c}_n)^2}{\max_{1 \leqslant i \leqslant n}(c_{ni} - \bar{c}_n)^2} = \infty. \tag{1.1}$$

For convenience the dependence of $\{X_{ni}\}$, $\{\varepsilon_{ni}\}$, and $\{c_{ni}\}$ on n will be suppressed, and the vectors $(c_1, c_2, \ldots, c_n)'$ and $(X_1, X_2, \ldots, X_n)'$ will be denoted by $\mathbf{c}$ and $\mathbf{X}$, respectively. We also define the quantities $B = \int_0^1 f(F^{-1}(t))\,dt$, $K = \int_0^1 f^2(F^{-1}(t))\,dt$ and $\sigma_c^2 = \sum (c_i - \bar{c})^2$ and let $d_i = c_i - \bar{c}$, $1 \leqslant i \leqslant n$.

In what follows let the vectors $\mathbf{c}$ and $\mathbf{X}$ be given. Also note that many of the statements made are true w.p.l, even though it is not stated explicitly. For each real Δ and for each t in $[0, 1]$ define

$$S(t, \Delta) := \sum_{i=1}^{n} d_i I(R_{ni\Delta} \leqslant nt)$$

where

$$R_{ni\Delta} := \sum_{j=1}^{n} I(X_j - \Delta c_j \leqslant X_i - \Delta c_i).$$

We may now write $M(\Delta) = \int_0^1 S^2(t, \Delta)\,dt$ and $W(\Delta) = \int_0^1 S(t, \Delta)\,dt$, where $M(\Delta)$ is the Cramér–von Mises statistic of the previous section and $\sigma_c^{-1} W(\Delta)$ is the Wilcoxon statistic [2].

For a fixed sample, $M(\Delta)$ is a step function (in Δ) whose points of discontinuity are contained in the set

$$\Gamma = \{(X_j - X_i)/(c_j - c_i);\ i < j \text{ and } c_i < c_j\}.$$

Set $\Delta_0 = \min\{\Delta; \Delta \in \Gamma\}$ and $\Delta_1 = \max\{\Delta: \Delta \in \Gamma\}$. Then for $c_i < c_j, \Delta < \Delta_0$ implies $\Delta < (X_j - X_i)/(c_j - c_i)$ and hence $R_{ni\Delta} < R_{nj\Delta}$. Thus the residuals $\{X_i - \Delta c_i;\ 1 \leqslant i \leqslant n\}$ are naturally ordered, except possibly within groups of residuals with equal c_i's. We may complete the ordering, without loss of generality, by appropriately renumbering and obtain

$$\begin{aligned} S(t, \Delta) &= \sum_{i=1}^{j} d_i \qquad \text{for } t \in \left[\frac{j}{n}, \frac{j+1}{n}\right),\ 1 \leqslant j \leqslant n-1 \\ &= 0 \qquad \text{for } t \in \left[0, \frac{1}{n}\right) \cup \{1\}. \end{aligned}$$

Hence $M(\Delta) = n^{-1} \sum_{j=1}^{n-1} [\sum_{i=1}^{j} d_i]^2 = M(\Delta_0-)$. When $\Delta > \Delta_1$ the ordering is reversed and $\sum_{i=1}^{n} d_i = 0$ implies $M(\Delta) = n^{-1} \sum_{j=1}^{n-1} [\sum_{i=1}^{j} d_i]^2 = M(\Delta_1+)$.

As Δ crosses Δ_0 only one pair of adjacent residuals cross. Let $c_k < c_m$ denote their respective regression constants. Then

$$M(\Delta_0-) - M(\Delta_0+) = n^{-1} \left\{ \left[\sum_{i=1}^{k} d_i \right]^2 - \left[d_m + \sum_{i=1}^{k-1} d_i \right]^2 \right\}.$$

Now $c_1 \leqslant c_2 \leqslant \cdots \leqslant c_n$ and $c_k < c_m$ imply $\sum_{i=1}^{k} d_i < d_m + \sum_{i=1}^{k-1} d_i \leqslant 0$. Thus $M(\Delta_0-) > M(\Delta_0+)$. Similarly, $M(\Delta_1+) > M(\Delta_1-)$. As a result, the following quantities are finite:

$$\beta^* = \min \left\{ s \in \Gamma; M(s+) = \inf_{\Delta \in \Gamma^c} M(\Delta) \right\},$$

$$\beta^{**} = \max \left\{ s \in \Gamma; M(s-) = \inf_{\Delta \in \Gamma^c} M(\Delta) \right\}.$$

We now define our estimator $\hat{\beta}$ for β by

$$\hat{\beta} = \tfrac{1}{2}(\beta^* + \beta^{**}).$$

NUMERICAL EXAMPLE. By the preceding remarks we may determine the value of $\hat{\beta}$ by identifying the set of slopes Γ and computing $M(\Delta-)$ for each $\Delta \in \Gamma$. Computation of $M(\Delta)$ is facilitated by using the formula

$$M(\Delta) = -n^{-1} \sum_{1 \leqslant i < j \leqslant n} d_i d_j |R_{ni\Delta} - R_{nj\Delta}|. \tag{1.2}$$

We consider here a two-sample problem; that is, we take $c_1 = c_2 = 0$ and $c_3 = c_4 = c_5 = c_6 = 1$ for the data

160, −26, 17, −150, −30, 12.

From Table I it is clear that $\beta^* = -148$, $\beta^{**} = -4$ and hence $\hat{\beta} = -76$.

TABLE I

Values of $M(\Delta-)$ for $\Delta \in \Gamma$

$\Delta \in \Gamma$	−310	−190	−148	−143	−124	−4	38	43
$M(\Delta-)$	0.6296	0.3519	0.1852	0.1296	0.1852	0.1296	0.1852	0.3519

2. Finite Sample Properties

(a) *Invariance*

A useful property of the estimator $\hat{\beta}$ is its translation invariance:

$$\hat{\beta}(\mathbf{X} + \gamma\mathbf{c}) = \hat{\beta}(\mathbf{X}) + \gamma \qquad \forall \text{ real } \gamma. \tag{2.1}$$

To verify (2.1) we note that $M(\Delta - \gamma)(\mathbf{X}) = M(\Delta)(\mathbf{X} + \gamma\mathbf{c})$ from the definition of M. Then $\beta^*(\mathbf{X} + \gamma\mathbf{c}) = \beta^*(\mathbf{X}) + \gamma$ and $\beta^{**}(\mathbf{X} + \gamma\mathbf{c}) = \beta^{**}(\mathbf{X}) + \gamma$ for all real γ, and (2.1) follows.

From (2.1) we conclude that

$$P_\beta(\hat{\beta} - \beta \leqslant z) = P_0(\hat{\beta} \leqslant z) \qquad \forall \text{ real } z \tag{2.2}$$

where P_β and P_0 indicate that the true parameter is assumed to be β and 0, respectively. Because of (2.2) we may assume hereafter that $\beta = 0$ without loss of generality.

(b) *Symmetry*

THEOREM 2.1. *$\hat{\beta}$ is symmetric about β if one of the following conditions holds:*

(i) *F is symmetric.*

(ii) $d_i = -d_{n-i+1}$, $1 \leqslant i \leqslant n$.

Proof of (i). Since for any real number a, $M(\Delta)(\mathbf{X}) = M(\Delta)(\mathbf{X} + a\mathbf{1})$, we may assume that F is symmetric about 0 without loss of generality. Now $\mathbf{X} \sim -\mathbf{X}$ so that $\hat{\beta}(\mathbf{X}) \sim \hat{\beta}(-\mathbf{X})$ and hence it suffices to show that $\hat{\beta}(-\mathbf{X}) \sim -\hat{\beta}(\mathbf{X})$. Let $\Theta = \{\Delta; X_j - X_i = \Delta(c_j - c_i)$, some $1 \leqslant i < j \leqslant n\}$. Then

$$R_{ni\Delta}(-\mathbf{X}) = n + 1 - R_{ni(-\Delta)}(\mathbf{X}), \qquad \Delta \in \Theta^C. \tag{2.3}$$

Using (1.2) and (2.3) and the fact that $\Theta = \Gamma$ w.p.l, we have (w.p.l) $M(-\Delta)(\mathbf{X}) = M(\Delta)(-\mathbf{X})$ for $\Delta \in \Gamma^C$. But this implies that $\hat{\beta}(-\mathbf{X}) = -\hat{\beta}(\mathbf{X})$, completing the proof of (i). ∎

Proof of (ii). Let $\mathbf{X}^*$ denote $(X_n, X_{n-1}, \ldots, X_1)'$. Since $\beta = 0$ we have $\mathbf{X} \sim \mathbf{X}^*$ and hence $\hat{\beta}(\mathbf{X}) \sim \hat{\beta}(\mathbf{X}^*)$. We show that $\hat{\beta}(\mathbf{X}^*) \sim -\hat{\beta}(\mathbf{X})$. But $R_{ni\Delta}(\mathbf{X}^*) = R_{n,n-i+1,-\Delta}(\mathbf{X})$ implies $S(t, \Delta)(\mathbf{X}^*) = -S(t, -\Delta)(\mathbf{X})$. Thus $M(\Delta)(\mathbf{X}^*) = M(-\Delta)(\mathbf{X})$ and $\hat{\beta}(\mathbf{X}^*) = -\hat{\beta}(\mathbf{X})$. Since $\mathbf{X}^* \sim \mathbf{X}$, we have $\hat{\beta}(\mathbf{X}^*) \sim -\hat{\beta}(\mathbf{X})$ and the proof of (ii) is completed. ∎

3. Asymptotic Distribution of $\sigma_c\hat{\beta}$

Throughout this section we retain the notation of Sections 0–2, the model of Section 0, and assumption (1.1).

Theorem 3.1. *Let* $0 < a < \infty$. *Then*

$$\sup_{\substack{0 \leqslant t \leqslant 1 \\ |\Delta| \leqslant a}} \sigma_c^{-1} |S(t, \Delta\sigma_c^{-1}) - S(t, 0) - \Delta\sigma_c f(F^{-1}(t))| \xrightarrow{P_0} 0. \tag{3.1}$$

Proof. Note that (1.7) of Koul [4] holds without assuming $\lim_{n\to\infty} \sup \sigma_c^{-1}\sqrt{n} \max_{1 \leqslant i \leqslant n} |d_i| < \infty$ [3], and that f is uniformly continuous by the assumptions of Section 0. The proof is completed using the techniques developed in Koul [4]. ∎

A consequence of Theorem 3.1 is

Lemma 3.1. *Let* $0 < a < \infty$ *and set* $T(\Delta) = \int_0^1 [S(t, 0) + \Delta\sigma_c f(F^{-1}(t))]^2 \, dt$. *Then*

$$\sup_{|\Delta| \leqslant a} \sigma_c^{-2} |M(\Delta\sigma_c^{-1}) - T(\Delta)| \xrightarrow{P_0} 0, \tag{3.2}$$

$$\sup_{|\Delta| \leqslant a} \sigma_c^{-2} \left| W^2(\Delta\sigma_c^{-1}) - \left(\int \left[S(t, 0) + \Delta\sigma_c f(F^{-1}(t))\right] dt\right)^2 \right| \xrightarrow{P_0} 0 \tag{3.3}$$

Proof. To establish (3.2) we note that $\sigma_c^{-1} \sup_{0 \leqslant t \leqslant 1} |S(t, 0)|$ is the Kolmogorov–Smirnov statistic which has a limiting distribution [2]. Thus (3.1) implies

$$\sup_{\substack{|\Delta| \leqslant a \\ 0 \leqslant t \leqslant 1}} \sigma_c^{-2} |S^2(t, \Delta\sigma_c^{-1}) - [S(t, 0) + \Delta\sigma_c f(F^{-1}(t))]^2| \xrightarrow{P_0} 0 \tag{3.4}$$

and (3.2) follows.

To establish (3.3) we note that

$$\sup_{|\Delta| \leqslant a} \sigma_c^{-1} \left| W(\Delta\sigma_c^{-1}) - \int_0^1 [S(t, 0) + \Delta\sigma_c f(F^{-1}(t))] \, dt \right| \xrightarrow{P_0} 0 \tag{3.5}$$

follows from (3.1). But $-\sigma_c^{-1} \int_0^1 S(t, 0) \, dt$ is the Wilcoxon statistic which has a limiting distribution [2]. Thus (3.3) follows from (3.5). ∎

LEMMA 3.2. *For every $\varepsilon > 0$ there exist positive real numbers N, a and d such that*

$$P_0\left[\sigma_c^{-2} W^2(0) < d, \inf_{|\Delta| = a} \sigma_c^{-2} W^2(\Delta\sigma_c^{-1}) \geqslant d\right] \geqslant 1 - \varepsilon \qquad \forall n \geqslant N.$$

Proof. Since $\sigma_c^{-1} W(0)$ has a limiting distribution, there exists a positive real number b such that $P_0[\sigma_c^{-1} |W(0)| \leqslant b] \geqslant 1 - \varepsilon \, \forall \, n$. If we also take $d > 2b^2$ and choose $a > [b + (3d/2)^{1/2}] B^{-1}$, we have

$$\sigma_c^{-2} W^2(0) \leqslant b^2 < d/2,$$

$$\inf_{|\Delta| = a} \sigma_c^{-2} \left(\int_0^1 [S(t, 0) + \Delta\sigma_c f(F^{-1}(t))] \, dt\right)^2 \geqslant 3d/2$$

on $\{\sigma_c^{-1} |W(0)| \leqslant b\}$. Choosing N according to (3.3) completes the proof. ∎

LEMMA 3.3. *For every $\varepsilon > 0$ and $d' > 0$ there exist positive real numbers a and N such that $n \geqslant N$ implies $P_0[\inf_{|\Delta| \geqslant a} \sigma_c^{-2} W^2(\Delta\sigma_c^{-1}) \geqslant d'] \geqslant 1 - \varepsilon$.*

Proof. In the proof of Lemma 3.2, take $d > \max\{d', 2b^2\}$. The proof is completed by using the fact that $W(\Delta)$ is nonincreasing in Δ [1, p. 35]. ∎

For the next lemma define $T_c(\cdot)$ as in V.3.5 of [2]. Also define the functional h, for a bounded continuous function y on $[0, 1]$, by $h(y) = K^{-1} \int_0^1 y(t) f(F^{-1}(t)) \, dt$ where $K = \int_{-\infty}^{\infty} f^3(x) \, dx$. Since h is continuous on $C[0, 1]$ and $\sup_{0 \leqslant t \leqslant 1} |T_c(t) + \sigma_c^{-1} S(t, 0)| \leqslant \sigma_c^{-1} \max_{1 \leqslant i \leqslant n} |d_i|$ we have $L_0(h(T_c)) \Rightarrow N(0, \sigma^2)$ where $\sigma^2 = K^{-2}[\int_0^1 G^2(t) \, dt - (\int_0^1 G(t) \, dt)^2]$ and $G(t) = \int_0^t f(F^{-1}(s)) \, ds$, $0 \leqslant t \leqslant 1$. Lemma 3.1 suggests the use of $\sigma_c \hat{\hat{\beta}} := h(\sigma_c^{-1} S(\cdot, 0))$ as an approximating statistic for $\sigma_c \hat{\beta}$. Its asymptotic distribution is given in

LEMMA 3.4. $L_0(\sigma_c \hat{\hat{\beta}}) \Rightarrow N(0, \sigma^2)$.

Proof. Use the above remarks and

$$|\sigma_c \hat{\hat{\beta}} - h(T_c)| \leqslant K^{-1} \|f\|_\infty \int_0^1 |\sigma_c^{-1} S(t, 0) + T_c(t)| \, dt$$

$$\leqslant (K\sigma_c)^{-1} \|f\|_\infty \max_{1 \leqslant i \leqslant n} |d_i|. \; ∎$$

LEMMA 3.5. *Given $\varepsilon > 0$ there exist positive real numbers a, b and N with $a > b$ such that $n \geqslant N$ implies $P_0[G_n(a, b)] \geqslant 1 - \varepsilon$ where*

$$G_n(a, b) := \left\{|\sigma_c \hat{\hat{\beta}}| \leqslant b, \inf_{|\Delta| \geqslant a\sigma_c^{-1}} M(\Delta) > \inf_{\substack{|\Delta| < a\sigma_c^{-1} \\ |\Delta| \notin \Gamma}} M(\Delta)\right\}.$$

Proof. Since $\sigma_c\hat{\hat{\beta}}$ and $\sigma_c^{-2}M(0)$ have limiting distributions there exists a positive real number b such that $P_0[\sigma_c|\hat{\hat{\beta}}| \leqslant b, \sigma_c^{-2}M(0) \leqslant b] \geqslant 1-\varepsilon/2$ for all n. Taking $d > b$ and noting that $M(\Delta) \geqslant W^2(\Delta)$ by the Cauchy–Schwarz inequality, it follows from Lemma 3.3 that there exist $a > b$ and $0 < N < \infty$ such that $P_0[\inf_{|\Delta| \geqslant a\sigma_c^{-1}} \sigma_c^{-2}M(\Delta) \geqslant d] \geqslant 1-\varepsilon/2$ for $n \geqslant N$. But then

$$\sigma_c^{-2}M(0) \geqslant \inf_{\substack{|\Delta| \leqslant a\sigma_c^{-1} \\ \Delta \notin \Gamma}} \sigma_c^{-2}M(\Delta) \Rightarrow$$

$$P_0[G_n(a,b)] \geqslant P_0 \left\{ \sigma_c|\hat{\hat{\beta}}| \leqslant b, \sigma_c^{-2}M(0) \leqslant b, \inf_{|\Delta| \geqslant a\sigma_c^{-1}} \sigma_c^{-2}M(\Delta) \geqslant d \right\}$$
$$\geqslant 1-\varepsilon. \quad \blacksquare$$

THEOREM 3.2. $L_0(\sigma_c\hat{\beta}) \Rightarrow N(0, \sigma^2)$.

Proof. We prove that $\sigma_c|\hat{\beta} - \hat{\hat{\beta}}| \to^{P_0} 0$. The theorem then follows from Lemma 3.4. Let $\varepsilon > 0$ and $\delta > 0$ be given. By Lemmas 3.1 and 3.6 there exist $a > b$ and $N > 0$ such that $n \geqslant N$ implies the following hold with probability at least $1-\varepsilon$:

(i) $\sigma_c^{-2}(M(\hat{\hat{\beta}} + \gamma\sigma_c^{-1}) - M(\hat{\hat{\beta}})) \geqslant \sigma_c^{-2}(T_c(\hat{\hat{\beta}}\sigma_c + \gamma) - T_c(\hat{\hat{\beta}}\sigma_c)) - \delta$
$= \gamma^2K - \delta \geqslant \delta$ whenever $|\gamma| \geqslant \sqrt{2\delta/K}$ and $\sigma_c|\hat{\hat{\beta}} - \gamma\sigma_c^{-1}| \leqslant a$,

(ii) $3\delta/4 \geqslant K\gamma^2 + \delta/2 = \sigma_c^{-2}(T_c(\hat{\hat{\beta}}\sigma_c + \gamma) - T_c(\hat{\hat{\beta}}\sigma_c)) + \delta/2$
$\geqslant \sigma_c^{-2}(M(\hat{\hat{\beta}} + \gamma\sigma_c^{-1}) - M(\hat{\hat{\beta}}))$ whenever $|\gamma| \leqslant \sqrt{\delta/4K}$, and

(iii) $\inf_{|\Delta| \geqslant \sigma_c^{-1}a} M(\Delta) > \inf_{|\Delta| < \sigma_c^{-1}a} M(\Delta)$.

Since (i)–(iii) imply that

$$P_0[\sigma_c|\hat{\hat{\beta}} - \beta^*| \leqslant \sqrt{2\delta/K}, \sigma_c|\hat{\hat{\beta}} - \beta^{**}| \leqslant 2\delta/K] \geqslant 1-\varepsilon \qquad \forall n \geqslant N,$$

and ε and δ can be made arbitrarily small, the theorem is proved. $\blacksquare$

4. ASYMPTOTIC VARIANCE OF $\sigma_c\hat{\beta}$

σ^2 can be computed from the formula given in Lemma 3.4 or from

$$\sigma^2 = K^{-2} \iint [F(x) \wedge F(y) - F(x)F(y)] f^2(x) f^2(y)\, dx dy. \tag{4.1}$$

The latter formula was used for $F = \Phi$. Table II indicates that $\hat{\beta}$ does quite well relative to the Wilcoxon, median, normal scores, and least-squares type estimators for β at underlying distributions with heavy tails.

TABLE II

Asymptotic Variances of Various Estimators for β

	σ_W^2	σ^2	σ_M^2	$\sigma_{\Phi^{-1}}^2$	σ_{LS}^2
D. Exp.	1.333	1.2	1	$\Pi/2 = 1.5707$	2
Logistic	3	3.0357	4	$\Pi = 3.1416$	$\Pi^2/3 = 3.2899$
Normal	$\Pi/3 = 1.0472$	1.0946	$\Pi/2 = 1.5707$	1	1
Cauchy	$\Pi^2/3 = 3.2899$	2.5733	$\Pi^2/4 = 2.4674$	4.65	—

Acknowledgments

This paper is based on the author's Ph.D. dissertation. The author wishes to thank Professor Hira L. Koul for his guidance in its preparation. Thanks are also due the referee for suggesting a shortened proof of Theorem 3.1.

References

[1] Hájek, J. (1969). *Nonparametric Statistics.* Holden-Day, San Francisco.

[2] Hájek, J. and Sidák, Z. (1967). *Theory of Rank Tests.* Academic Press, New York.

[3] Koul, H. (1970). Some convergence theorems for ranks and weighted empirical cumulatives. *Ann. Math. Statist.* **41** 1768–1773.

[4] Koul, H. (1977). Behavior of robust estimators in the regression model with dependent errors. *Ann. Statist.* **5** 681–699.